C:\WINDOWS\TEMP\AirToxics.wpd

Date 2000

EPA-SAB-EC-00-XXX

Honorable Carol M. Browner
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Ave.
Washington, D.C. 20460

RE: Review of draft Air Toxics Monitoring Strategy Concept Paper

Dear Ms. Browner:

On March 29-30, 2000, the Science Advisory Board's (SAB's) Air Toxics Monitoring Subcommittee of the SAB Executive Committee reviewed the February 29, 2000 drafts of the Air Toxics Monitoring Strategy Concept Paper and the Protocol for Model-to-Monitor Comparisons. The Office of Air Quality, Planning, and Standards (OAQPS) prepared both documents as part of the National Air Toxics Assessment (NATA). The accompanying SAB report responds to charge questions concerning the air toxic monitoring objectives and principles, the phased strategy for the design of a national air toxics network, and the Model-to-Monitor evaluation protocol.

In briefest terms, the Agency is taking a sound, scientific approach with the available resources. OAQPS has appropriately decided to address a limited number of objectives. They are approaching these objectives in a logical, informed and step-wise fashion. The Subcommittee expects that systematic planning will continue to be done as strategies are phased in to allow optimal use of available resources. It is crucially important not to spread the available resources so thinly that nothing can be done well. Therefore, the Subcommittee's additional suggestions for valuable work should be

considered in the event that additional resources become available. These suggestions are included in this report.

In summary, the Subcommittee finds that the concept paper presents a reasonable phased strategy to design a national air toxics network. The Agency has identified the most important uses for ambient air toxics data and the Subcommittee offered suggestions for other types of air toxics monitoring data or data uses. As first steps in the design of the national network, the Subcommittee endorses the Agency's data analysis of existing air toxics monitoring data, coupled with focused pilot studies for a core set of hazardous air pollutants (HAPs) in a limited number of areas. Overall, the Subcommittee supports the goals of the Model-to-Monitor evaluation protocol and has provided a number of specific comments for applications of the protocol.

The Subcommittee wishes to highlight certain findings concerning the existing documents:

a. The Subcommittee commends the Agency for the quality of both the Air Toxics Monitoring Concept Paper and the Model-to-Monitor strategy. A great deal of careful thought and evaluation has gone into this process. Both reports are terse, clear, and well written.

b. Although the Agency has identified the most important uses for ambient air toxics data, the Subcommittee identified additional types of air toxics monitoring data and data uses.

c. The concept paper presents a reasonable phased strategy to design a national air toxics network.

d. With only minor modifications, the Agency's strategy for neighborhood-scale sampling over a 24-hour period is appropriate.

e. The Agency's data analysis of existing air toxics monitoring data, coupled with focused pilot studies for a core set of hazardous air pollutants (HAPs) in a limited number of areas, are appropriate first steps in the design of the national network.

f. The Subcommittee encourages the use of tools, such as the data quality objective process, to improve the relevance and reliability of the exposure information required by the Agency.

g. Overall, the Subcommittee supports the goals of the Model-to-Monitor evaluation protocol and provides a number of specific comments for improving the application of the protocol.

The Subcommittee also commented on the possibilities for an expanded study because there are a number of areas in which the monitoring activities and data evaluation could be improved or expanded, if more resources are made available.

a. A well-defined and consistent scientific framework addressing both modeling and measurement problems for hazardous air pollutants would be valuable.

b. It would be useful to understand how well monitoring and modeling define source-to-concentration relationships for certain types of pollutants.

c. It is not clear how outdoor ambient air measurements are to be related to population exposure which occurs mostly indoors. Consideration should be given to parallel pilot studies on indoor air monitoring, using techniques having similar detection levels as those in the proposed outdoor study, in order to establish indoor/outdoor relationships that would permit better estimation of indoor exposures and the apportionment of the outdoor contributions to the exposures.

d. If multimedia pollutants were included in both the monitoring and modeling framework, additional exposure routes could be considered. This consideration is important because several classes of HAPs are multimedia pollutants; for example, metals and semivolatile organic compounds that are transferred through food chains, and for which the inhalation route plays a relatively minor role.

In the future, NATA and the monitoring data it provides will be a resource of enormous scientific value. Understanding air toxics in the environment is an important area where much has been achieved and yet much remains to be done. Because currently the Agency is rich in models and poor in data, data collection efforts such as this should be given high priority. Therefore, the Subcommittee hopes that the Congress and the Agency will continue to provide the resources to OAQPS to expand this program and that the Agency continues to collaborate with state and local agencies.

We appreciate the opportunity to provide advice on this effort at this early stage. The Agency staff was open, collegial, and cognizant of capabilities and limitations of the concept paper. They were also accepting of and responsive to the Subcommittee's suggestions. The SAB is open to reviewing elements of this monitoring strategy at appropriate stages of its development. We look forward to the Agency's response from the Office of Air and Radiation.

Sincerely,

Dr. Morton Lippmann
Interim Chair
Executive Committee

Dr. Thomas McKone
Chair
Air Toxics Monitoring Subcommittee

1

NOTICE

This report has been written as part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

Distribution and Availability: This Science Advisory Board report is provided to the EPA Administrator, senior Agency management, appropriate program staff, interested members of the public, and is posted on the SAB website (www.epa.gov/sab). Information on its availability is also provided in the SAB's monthly newsletter (*Happenings at the Science Advisory Board*). Additional copies and further information are available from the SAB Staff.

ABSTRACT

On March 29-30, 2000, the Science Advisory Board's (SAB's) Air Toxics Monitoring Subcommittee of the SAB Executive Committee conducted a peer review of the February 29, 2000 drafts of the Air Toxics Monitoring Strategy Concept Paper and the Protocol for Model-to-Monitor Comparisons. Both documents are part of the National Air Toxics Assessment (NATA) and were prepared by the Office of Air Quality, Planning, and Standards (OAQPS).

The Subcommittee commended the Agency for their effort in developing both the Air Toxics Monitoring Concept Paper and the Monitoring-to-Models strategy. The concept paper presents a reasonable phased strategy to design a national air toxics network. The Subcommittee endorses the Agency's data analysis of existing air toxics monitoring data, coupled with focused pilot studies for a core set of hazardous air pollutants (HAPs) in a limited number of areas. The Agency has identified the most important uses for ambient air toxics data and the Subcommittee offered suggestions for other types of air toxics monitoring data and/or data uses.

NATA and the monitoring data it provides will be a resource of enormous scientific value. The Subcommittee identified a number of areas in which the monitoring activities and data evaluation could be improved or expanded, particularly if more resources are made available.

Keywords: hazardous air pollutants, air toxics, monitoring, NATA.

U.S. ENVIRONMENTAL PROTECTION AGENCY
SCIENCE ADVISORY BOARD
AD HOC SUBCOMMITTEE OF THE EXECUTIVE COMMITTEE:
AIR TOXICS MONITORING STRATEGY SUBCOMMITTEE
FY00

CHAIRMAN*

Dr. Thomas E. McKone, Staff Scientist Lawrence Berkely Laboratory and Adjunct Professor, School of Public Health, University of California, Berkeley, CA

MEMBERS OF SAB COMMITTEES*

Dr. Maria Morandi, Assistant Professor, University of Texas Health Science Center at Houston, School of Public Health, Houston, TX

CONSULTANTS TO THE SAB*

Dr. Rebecca Ann Efroymson, Research Staff Member, Environmental Sciences Division, Oak Ridge National Laboratory,Oak Ridge, TN

Dr. Timothy Larson, Professor, Dept. of Civil and Environmental Engineering, University of Washington, Seattle, Washington

Dr. Karl R. Loos, Environmental Advisor, Equilon Enterprises LLC, Houston, Texas

Dr. Jana Milford, Associate Professor, Department of Mechanical Engineering, University of Colorado, Boulder, CO

Mr. Thomas Natan, National Environmental Trust,Washington, DC

Dr. Edo D. Pellizzari, Research Triangle Institute, P.O. Box 12194, 3040 Cornwallis Road,Research Triangle Park, NC 27709

Mr. Doug Splitstone, Splitstone & Assoc. Murrysville, PA

Mr. Mel Zeldin, Director, Monitoring and Analysis Division, South Coast Air Quality Management District, Diamond Bar, CA

SCIENCE ADVISORY BOARD STAFF

Kathleen White Conway, Designated Federal Officer, Science Advisory Board (1400), Environmental Protection Agency, 1200 Pennsylvania Avenue, NW, Washington, DC 20460

Mary L. Winston, Management Assistant, Science Advisory Board (1400), Environmental Protection Agency, 1200 Pennsylvania Avenue, NW, Washington, DC

* Members of this SAB Subcommittee consist of

a. SAB Members: Experts appointed by the Administrator to two-year terms to serve on one of the 10 SAB Standing Committees.

b. SAB Consultants: Experts appointed by the SAB Staff Director to a one-year renewable term to address a particular issue.

U.S. ENVIRONMENTAL PROTECTION AGENCY
SCIENCE ADVISORY BOARD
EXECUTIVE COMMITTEE
FY-2000

INTERIM CHAIR

Dr. Morton Lippmann, Professor, Nelson Institute of Environmental Medicine, New York University School of Medicine, 57 Old Forge Road, Tuxedo, NY 10987

MEMBERS

Dr. Henry A. Anderson, Chief Medical Officer, Wisconsin Division of Public Health, 1414 East Washington Avenue, Room 96, Madison, WI 53703

Dr. Richard J. Bull, MoBull Consulting, Inc., 8382 Gage Boulevard, Suite O, Box No. 511, Kennewick, WA 99336

Dr. Maureen L. Cropper, Principal Economist, DECRG, The World Bank, 1818 H Street, NW, MC2-521, Washington, DC 20433

Dr. Kenneth W. Cummins, Senior Advisory Scientist, California Cooperative Fishery Research Unit and Adjunct Professor, Fisheries Department, Humboldt State University, Arcata, CA 95521

Dr. Linda Greer, Senior Scientist, Natural Resources Defense Council, 1200 New York Avenue, NW, Suite 400, Washington, DC 20005

Dr. Hilary I. Inyang, University Professor and Director, Center for Environmental Engineering, Science and Technology (CEEST), University of Massachusetts Lowell, One University Avenue, E-114, Lowell, MA 01854

Dr. Janet A. Johnson, Senior Radiation Scientist, Shepherd Miller, Inc., 3801 Automation Way, Suite 100, Fort Collins, CO 80525

Dr. Roger E. Kasperson, University Professor and Director, The George Perkins Marsh Institute, Clark University, 950 Main Street, Worcester, MA 01610-1477

Dr. Joe L. Mauderly, Director & Senior Scientist, Lovelace Respiratory Research Institute, P.O. Box, Albuquerque, NM 87185

Dr. M. Granger Morgan, Head, Department of Engineering & Public Policy, Carnegie Mellon University, 129 Banker Hall, Frew Street, Pittsburgh, PA 15213

Dr. William Randall Seeker, Senior Vice President, General Electric Energy and Environmental Research Corp., 18 Mason, Irvine, CA 92618

Dr. William H. Smith, Professor of Forest Biology, Yale University, 205 Prospect Street, New Haven, CT 06511

Dr. Robert N. Stavins, Albert Pratt Professor of Business and Government, Faculty Chair, Environment and Natural Resources Program, John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138

Dr. Mark J. Utell, Professor of Medicine and Environmental Medicine, University of Rochester Medical Center, Pulmonary Unit - Box 692, 601 Elmwood Avenue, Rochester, NY 14642-8692

Dr. Terry F. Young, Senior Consulting Scientist, Environmental Defense Fund, 5655 College Avenue, Oakland, CA 94611

LIAISON FOR CHILDRENS HEALTH PROTECTION ADVISORY COMMITTEE

Mr. J. Thomas Carrato, Assistant General Counsel, Regulatory Affairs, Monsanto Company, Mail Zone C2SA, 800 N. Linbergh Boulevard, St. Louis, MO 63167

LIAISON FOR SCIENCE ADVISORY PANEL

Dr. Ronald Kendall, Director & Professor, The Institute of Environmental & Human Health, Texas Tech University/Texas Tech University Health Sciences Center, 1207 Gilbert Drive, Reese Center, Building #555, Lubbock, TX 79416

LIAISON FOR ORD BOARD OF SCIENTIFIC COUNSELORS

Dr. Costel D. Denson, Professor of Chemical Engineering, University of Delaware, 1184 Corner Ketch Road, Newark, DE 19711

SCIENCE ADVISORY BOARD STAFF

Dr. Donald G. Barnes, Staff Director/Designated Federal Officer, Environmental Protection Agency, Science Advisory Board (1400A),1200 Pennsylvania Avenue, NW, Ariel Rios North Lobby, Room 6450, Washington, DC 20460, (202) 564-4533; FAX (202) 501-0323 (barnes.don@epa.gov) (FedEx: Washington, DC 20004)

Ms. Priscilla Y. Tillery-Gadson, Program Specialist, Environmental Protection Agency, Science Advisory Board (1400A), 1200 Pennsylvania Avenue, NW, Ariel Rios North Lobby, Room 6450, Washington, DC 20460, (202) 564-4543; FAX (202) 501-0323 (tillery.priscilla@epa.gov) (FedEx: Washington, DC 20004)

Ms. Betty B. Fortune, Office Assistant, Environmental Protection Agency, Science Advisory Board (1400A), 1200 Pennsylvania Avenue, NW, Ariel Rios North Lobby, Room 6450, Washington, DC 20460, (202) 564-4534; FAX (202) 501-0323 (fortune.betty@epa.gov) (FedEx: Washington, DC 20004)

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY 1

2 INTRODUCTION 3
2.1 Background 3
2.2 Charge 5
2.3 SAB Review Process 7

3 Responses to the Charge 8
3.1 Summary of the Subcommittee Responses 9
3.1.1 Question 1 9
3.1.2 Question 2 12
3.1.3 Question 3. 14
3.2 Detailed Responses to Question 1 16
3.2.1 Question 1(a) 18
3.2.2 Question 1(b) 20
3.1.3 Question 1(c) 22
3.3 Detailed Responses to Question 2 24
3.3.1 Question 2(a) 25
3.3.2 Question 2(b 26
3.4 Detailed Responses to Question 3 on Model-to-Monitor Comparison 29
3.4.1 Question 3(a) 29
3.4.2 Question 3(b) 34
3.4.3 Question 3(c) 34

GLOSSARY 41

REFERENCES 42

APPENDIX A: Exposure Measurement Issues

APPENDIX B: Summary of Elements of the EPA Quality System and an Introduction to the Data Quality Objectives Process

1 EXECUTIVE SUMMARY

On March 29-30, 2000, the Science Advisory Board's (SAB's) Air Toxics Monitoring Subcommittee of the SAB Executive Committee reviewed the February 29, 2000 drafts of the Air Toxics Monitoring Strategy Concept Paper and the Protocol for Model-to-Monitor Comparisons. The Office of Air Quality, Planning, and Standards (OAQPS) prepared both documents in support of the National Air Toxics Assessment (NATA) program.

The Subcommittee commends the Agency for the documents. A great deal of careful thought and evaluation has gone into this process with the result that both reports are terse, clear, and well-written.

The Agency has identified the most important uses for ambient air toxics data, and the concept paper presents a reasonable phased strategy to design a national air toxics network. The Subcommittee endorses the Agency's data analysis of existing air toxics monitoring data and focused pilot monitoring studies for a core set of hazardous air pollutants (HAPs) in a limited number of areas as first steps in the design of the national network. With minor changes, the Subcommittee supports the goals of the Model-to-Monitor evaluation protocol and the Agency's strategy for neighborhood-scale sampling over 24-hour periods.

The NATA and the monitoring data it provides will be a resource of enormous scientific value for understanding air toxics in the environment. This is an important area where much has been achieved and yet much remains to be done. Because currently the Agency is rich in models and poor in data, data collection efforts such as this are very important. Therefore, the Subcommittee identified a number of areas where the

monitoring activities and data evaluation could be improved or expanded, particularly if more resources are made available.

One Subcommittee concern is the initial exclusion of multimedia pollutants from both the monitoring and modeling framework and the related factor that only inhalation is considered as an exposure route. Even though multimedia pollutants are not the first priority of an air-monitoring program, several important classes of HAPs are multimedia pollutants. Metals and semivolatile organic compounds, for example, are transferred through food chains; for these, inhalation is not the primary direct route of exposure. By excluding such compounds from the first phase of the program, there is some chance they could be excluded over the long term because absence of information could be interpreted as the absence of a problem. Thus, the Subcommittee recommends a process that provides a continuing incentive for including multimedia and multipathway concentration and exposure data.

The Subcommittee also recommends the development and use of tools, such as the Data Quality Objective Process, to link sampling strategies to the relevance and reliability of the exposure information required by the Agency.

The Subcommittee hopes that the Congress and the Agency will continue to provide the resources to OAQPS to expand this program. It is also important to continue the Agency's collaborations with state and local agencies.

2 INTRODUCTION

2.1 Background

The Clean Air Act (CAA) regulates 188 hazardous air pollutants (HAPs), also called "air toxics" because they have been associated with a wide variety of adverse health effects. These air toxics are emitted from multiple sources and result in population exposure. Typically, people experience exposures to multiple HAPs from many sources. Exposures of concern result not only from the inhalation of these HAPs, but also, for some HAPs, from multi-pathway exposures to air emissions. For example, air emissions of mercury are deposited in water and people are exposed to mercury through their consumption of contaminated fish.

One of EPA's goals is to reduce air toxics emissions by 75% from 1993 levels. When tools are available to assess the residual risk, EPA plans to modify that goal to focus on reducing risks associated with exposure to air toxics. EPA's long-term goal is to eliminate unacceptable risks of cancer and other significant health problems resulting from exposures to air toxics emissions and to substantially reduce or eliminate adverse effects on our natural environment.

To meet these goals, EPA has developed an Air Toxics Program (ATP) to characterize, prioritize, and equitably address the impacts of HAPs on the public health and the environment. The ATP seeks to address air toxics problems through a strategic combination of agencies, activities and authorities, including regulatory approaches and voluntary partnerships. It includes four elements:

a. Source-specific standards and sector-based standards, including Section 112 standards; i.e. Maximum Achievable Control Technology (MACT), Generally Achievable Control Technology (GACT), residual risk standards, and Section 129 standards.

b. National, regional, and community-based initiatives to focus on multi-media and cumulative risks, such as the Integrated Urban Air Toxics Strategy, Great Waters, Mercury initiatives, Persistent Bioaccumulative Toxics (PBT) and Total Maximum Daily Load (TMDL) initiatives, and Clean Air Partnerships.

c. National Air Toxics Assessment (NATA) that will help EPA identify areas of concern, characterize risks, and track progress. These activities include expanded air toxics monitoring, improving and periodically updating emissions inventories, national-and local-scale air quality and exposure modeling, and continued research on effects and assessment tools, leading to improved characterizations of air toxics risk and reductions in

risk resulting from ongoing and future implementation of air toxics emissions control standards and initiatives.

d. Education and outreach.

The ATP depends on quantifying the impacts of air toxics emissions on public health and the environment. Therefore, EPA has initiated a National Air Toxics Assessment (NATA) to provide the best technical information regarding air toxics emissions, ambient concentrations, and health impacts to support the development of sound policies in the ATP. These activities include:

a. measurement of air toxics emission rates from individual pollution sources;

b. compilation of comprehensive air toxics emission inventories for local, State, and national domains;

c. measurement of ambient concentrations of air toxics at monitoring sites throughout the nation;

d. analyses of patterns and trends in ambient air toxics measurements;

e. estimation of ambient and multimedia air toxics concentrations from emission inventories, using dispersion and deposition modeling;

f. estimation of human and environmental exposures to air toxics;

g. assessment of risks due to air toxics; and

h. ongoing research in the above areas to improve assessments over time.

Emissions data, ambient concentration measurements, modeled estimates, and health and environmental impact information are all needed to fully assess air toxics impacts and to characterize risk. Specifically, emissions data are needed to quantify the sources of air toxics and aid in the development of control strategies. Ambient monitoring data can be used to evaluate the atmospheric dispersion and deposition which describe the fate and transport of air toxics in the atmosphere. Because ambient measurements cannot be made everywhere, modeled estimates are used to extrapolate to locations without monitors. A combination of reliable modeling systems along with a well-designed ambient network is thought to be the best approach for estimating ambient concentrations and population exposure across the nation.

Exposure assessments and health effects information integrate all of these data into an understanding of the implications of air toxics impacts and to characterize air toxics risks. Ambient measurements provided by routine monitoring programs together with personal exposure measurements obtained from ongoing research studies are important to evaluate these air quality and exposure models.

The Air Toxics Monitoring Strategy Concept Paper and related documents were drafted and submitted for review to provide a logical and scientifically strong basis to meet these data needs.

2.2 Charge

The focus of the present SAB review was to evaluate the adequacy of the February 29, 2000, drafts of the Air Toxics Monitoring Strategy Concept Paper and the Protocol for Model-to-Monitor Comparisons. The former describes a phased approach towards meeting the monitoring objectives of the Air Toxics Program. Both documents are part of the National Air Toxics Assessment and were prepared by the Office of Air Quality, Planning, and Standards. The Agency asked the SAB to focus on three specific questions.

"1. Does the air toxics monitoring concept paper describe appropriate air toxic monitoring objectives and principles, particularly ones that will permit the collection of monitoring data to support the initial National Air Toxics Assessment activities? Specifically,

(a) Does the Subcommittee believe that we identified the most important uses for ambient air toxics data? Are there other types of air toxics monitoring data or data uses that should also be identified for near-term or for future air toxics monitoring activities?

(b) Does the Subcommittee believe that neighborhood-scale monitoring is appropriate for evaluating ASPEN air quality predictions and later for developing long term ambient air quality trends? Are there other appropriate monitoring scales, perhaps for other data uses, that the Subcommittee would suggest?

(c) Does the Subcommittee believe that a basic 24-hour sample taken at a frequency sufficient to fulfill the objectives of the program is adequate to provide the model reality check and supply data for the characterization of ambient hazardous air pollutant (HAPs) concentrations? In particular, what are the Subcommittee's thoughts on the use of 24-hour samples collected once in 12 days for model evaluation and at a more frequent (say, 1-in-6 or 1-in-3 day) schedule in the future for trends assessment at permanently located monitoring sites?"

"2. Does the air toxics monitoring concept paper present a reasonable phased strategy to design a national air toxics network?

(a) Does the Subcommittee believe that data analyses of existing air toxics monitoring data, coupled with focused pilot studies for a core set of HAPs in a limited number of areas are appropriate first steps in the design of the national network. What additional or alternative approaches are suggested?

(b) Given that the State and local agencies have been measuring ambient air toxics with the Toxic Organic (TO) and Inorganic (IO) methods as described in the paper, are these methods appropriate for the continued routine monitoring of the target Urban Air Toxics Strategy compounds in a national monitoring network? If not, are there alternative methods which the Subcommittee would recommend?"

"3. In addition to your comments on the overall monitoring strategy, we seek your advice on the monitor-to-model evaluation protocol.

(a) Do the data analysis approaches provide enough information to allow appropriate interpretations of model results to support the development of model improvements in the future and to assist with the design of the national monitoring network?

(b) Are there some HAPs for which these approaches appear inadequate? If so, can the Subcommittee suggest alternative approaches for these?

(c) As noted in the paper, annual-average concentrations and comparisons to modeled estimates can be uncertain when a large percentage of the measurements are below the method detection limit (MDL). To estimate annual-average concentrations from monitoring data, EPA generally substitutes one-half the MDL. Does the Subcommittee suggest any alternative statistical approaches?"

2.3 SAB Review Process

Members of the Subcommittee were recruited from a variety of SAB Standing Committees and consultants to form a body of reviewers well-acquainted with the work of the SAB's Clean Air Scientific Advisory Committee, the Drinking Water Committee, the Environmental Engineering Committee, the Environmental Modeling Subcommittee, the Integrated Human Exposure Committee, and the Research Strategies Advisory Committee. Various Subcommittee members also served on the key relevant National Research Council reviews. The purpose of this formation was to provide continuity and consistency in advice on this important topic.

The Subcommittee met in public session on March 29-30, 2000, in Washington, DC. This report is based upon written comments prepared before and during the meeting by Subcommittee members and subsequently edited by the Subcommittee and approved by mail May 30, 2000. The report was then transmitted to the SAB's Executive Committee, for approval at a public meeting **[date]**

3 Responses to the Charge

Controlling the exposure of human populations and ecosystems to environmental contaminants depends upon understanding the links between multiple pollutant sources, exposure pathways, and adverse effects. This is a large and difficult problem. Both researchers and regulators are aware that air toxics data needs are real and pressing. Although much has been learned, much remains to be done. Currently, the Agency is rich in models and poor in data. Therefore, NATA and the monitoring data it provides will be a resource of enormous scientific value.

Given this size and complexity of the problem and the paucity of the data, it is not surprising that the Agency documents contain an ambitious list of goals for the proposed monitoring system and that the Subcommittee has identified even more. The challenge is in selecting from among these important needs a set of goals which are achievable within the resources available.

In this chapter, the Subcommittee, based on members' expertise and experience, provides guidance on which goals to pursue in which order. The Subcommittee did not make a detailed consideration of resources. The use of the Agency's Data Quality Objective Process (See Appendix B.) or other systematic planning process can help managers select among the many worthy and competing goals because it links the decision(s) to be made, the data and certainty required, and the project-specific blueprint for obtaining data appropriate for decision-making. Such planning processes provide a reality check. If the goal cannot be achieved within the resources available, then another (achievable) goal may be preferable.

In summary, the Subcommittee finds that the concept paper presents a reasonable phased strategy to design a national air toxics network. The Agency has identified the most important uses for ambient air toxics data and the Subcommittee offered suggestions for other types of air toxics monitoring data or data uses. As first steps in the design of the national network, the Subcommittee endorses the Agency's data analysis of existing air toxics monitoring data, coupled with focused pilot studies for a core set of hazardous air pollutants (HAPs) in a limited number of areas. Overall, the Subcommittee supports the goals of the Model-to-Monitor evaluation protocol and has provided a number of specific comments for applications of the protocol.

The Subcommittee identified a number of areas in which the data evaluation could be improved or expanded, if more resources are made available; these include indoor monitoring and multimedia studies. The proposed program does not address actual

population exposures, which mostly occur indoors. Current personal exposure monitoring methods do not have the sensitivity of those used in the proposed network. A pilot program to measure indoor levels with similar methods should be considered along with methods to determine the relationship of outdoor concentrations to indoor levels.

For many pollutants, transport through air is governed by deposition to and re-emission from soil, water and vegetation. These are difficult to measure in air. For the purposes of this report, these are called multi-media pollutants. EPA's proposed approach excludes multimedia pollutants from both the monitoring and modeling framework, and only inhalation is considered as an exposure route. Although multimedia pollutants are not the first priority of an air-monitoring program, several important classes of HAPs are multimedia pollutants; for example metals and semivolatile organic compounds that are transferred through food chains. The Subcommittee is not concerned about this priority, but fears that by excluding such compounds from the first phase of the program, there is some chance they could be excluded over the long term.

Because data collection efforts such as this should be given high priority, the Subcommittee hopes that the Congress and the Agency will continue to provide the resources to OAQPS to expand this program.

The following section begins with a summary of the Subcommittee's response and recommendations for each of the questions. More detailed discussions to support and expand on the recommendations follow the summaries.

3.1 Summary of the Subcommittee Responses

3.1.1 Question 1

Does the air toxics monitoring concept paper describe appropriate air toxic monitoring objectives and principles, particularly ones that will permit the collection of monitoring data to support the initial National Air Toxics Assessment activities?

The Subcommittee found that the monitoring objectives and principles described in the concept paper are appropriate and will permit the collection of the data necessary to support the initial NATA activities. However, the Subcommittee recommends that to better meet their specific aims, the EPA should develop a detailed, step-wise process that will make clear how they will achieve the primary objectives of the program. In this process, careful consideration should be given to the following issues:

a. Prioritize activities and milestones needed to achieve the primary objectives of NATA, following the Government Performance and Results Act (GPRA) model.

b. Develop a set of criteria for judging the adequacy of the network in terms of the objectives.

c. Develop data quality criteria for each of the activities.

d. Establish acceptable levels of data uncertainties (measurement errors) and model estimate uncertainties for both extant data and the new data to be collected.

e. Develop a process for including potential modification(s) of the plan in order to address unforeseen events or new findings.

f. As much as possible, involve the regional offices and local/state authorities that will be expected to contribute to NATA activities form the initial stages of plan development.

g. The revised Concept Paper discuss guidelines for siting monitors in rural areas; and

3.1.1.1 Question 1(a)

Does the Subcommittee believe that the EPA has identified the most important uses for ambient air toxics data? Are there other types of air toxics monitoring data or data uses that should also be identified for near-term or for future air toxics monitoring activities?

The Subcommittee believes that the document identifies the most important uses of the ambient air toxics data as they relate to the main objectives of NATA. As with any other data collection effort, there will be unforeseen, future uses of these data. The Subcommittee identified the following near-term and future uses of these data:

a. Identification of unpermitted and unreported emissions.

b. Input to permitting and siting decisions for new facilities.

c. "Reality check" on actual emission reductions vs. model-derived estimates.

d. Support and evaluation of the impact of other EPA and local/state programs.

e. Evaluation tool for other models that include or require air toxics information, including multi-media and cross-media exposure models.

However, the Subcommittee notes that some of these potential uses could contradict Agency policy as specified in 40 CFR Part 51 of April 21, 2000, section 10.2.2 and thus could be inappropriate uses of current source-receptor modeling techniques.

Because of the foreseen and unforeseen uses of these data, the Subcommittee recommends that:

a. input be obtained from other Agency programs that deal with multiple uses of data, including the Quality System (See Appendix B for description) and the efforts on secondary uses of data.

b. EPA develop a careful description of the data collected as well as a description of the capabilities and limitations of the data.

c. Utilize the Pilot Study to gain needed information on appropriate samping frequency and time-resolved (for example, day/night) sampling.

d. EPA consider giving to multimedia monitoring for persistent air toxics, given the potential application as an evaluation tool for other models that include or require air toxics information.

e. EPA collaborate with other on-going exposure and health surveys as a means on increasing the potential uses of the data.

3.1.1.2 Question 1(b)

Does the Subcommittee believe that neighborhood-scale monitoring is appropriate for evaluating ASPEN air quality predictions and later for developing long term ambient air quality trends? Are there other appropriate monitoring scales, perhaps for other data uses, that the Subcommittee would suggest?

The Subcommittee found that the neighborhood-scale monitoring approach is generally appropriate for the evaluation of ASPEN estimates and long-term air quality trends. However, because of the national scope of the program and the diversity of sites, this scale may not be applicable in all cases and may need to be modified. The Subcommittee recommends that:

a. Neighborhood-scale sampling should remain the main focus of the strategy. However, over the longer term, the monitoring scale used for toxic air pollutants should be guided by both the objectives of the monitoring program and the characteristics of the pollutants being monitored

b. The Agency should consider a micro-scale-type emission site in the Pilot study to assess variability within a neighborhood scale. But though it can be important, microscale sampling should be limited because of the resources needed; this situation may be more efficiently addressed through modeling in the main phase of the program.

c. Assuring that modern laboratory equipment (in some cases less than five years old) be used for analysis of samples.

d. Ecology-effects oriented siting should be considered, in addition to the human exposure focus.

e. Co-locating monitors at existing monitoring locations used for other purposes should be evaluated carefully in the context of the objectives of NATA.

3.1.1.3 Question 1(c)

Does the Subcommittee believe that a basic 24-hour sample taken at a frequency sufficient to fulfill the objectives of the program is adequate to provide the model reality check and supply data for the characterization of ambient hazardous air pollutant (HAPs) concentrations? In particular, what are the Subcommittee's thoughts on the use of 24-hour samples collected once in 12 days for model evaluation and at a more frequent (say, 1-in-6 or 1-in-3 day) schedule in the future for trends assessment at permanently located monitoring sites?

The Subcommittee believes that the one in twelve day, 24-hour sampling frame, while reasonable in principle, may not be consistent with the objectives of NATA for all compounds and foreseen uses of the data. As with the issue of spatial location of sites, the development of data quality criteria should assist in this process. The Subcommittee recommends that EPA:

a. Develop a multi-tier sampling frame after careful consideration of the objectives of NATA, the data quality criteria, the nature of the air toxic (or class of compound) and the temporal variability of the emission sources, and the specific uses of the data.

b. Consider also future uses of the data in developing the temporal sampling time frame.

3.1.2 Question 2

Does the air toxics monitoring concept paper present a reasonable phased strategy to design a national air toxics network?

3.1.2.1 Question 2(a)

Does the Subcommittee believe that data analyses of existing air toxics

monitoring data, coupled with focused pilot studies for a core set of HAPs in a limited number of areas are appropriate first steps in the design of the national network. What additional or alternative approaches are suggested?

The Subcommittee determined that the proposed phased strategy for NATA is reasonable as an initial approach to developing the network. However, the Subcommittee suggested alternate approaches for evaluating the information derived from the program. The Subcommittee noted that there should be careful definition of the "decision unit" to which the samples are applied. In addition, the subcommittee observed that there is a significant leap between ambient air concentration measurements and estimation of exposures, which requires also indoor concentration data and information on time-activity patterns. Since these data will not be collected, the Subcommittee recommends that EPA consider the need for developing penetration factors in parallel research programs. (Penetration factors help establish the relationship between indoor and outdoor air quality.)

3.1.2.2 Question 2(b)

Given that the State and local agencies have been measuring ambient air toxics with the Toxic Organic (TO) and Inorganic (IO) methods as described in the paper, are these methods appropriate for the continued routine monitoring of the target Urban Air Toxics Strategy compounds in a national monitoring network? If not, are there alternative methods that the Subcommittee would recommend?

The Subcommittee agrees with EPA that the TO and IO methods are generally appropriate for use in the national monitoring network. However, the Subcommittee noted some limitations to these methods. Some of these limitations are compound-specific, and others relate to the resources required to meet the objectives and stated sampling goals. The Subcommittee recommends that the Agency consider:

a. Evaluating each method in light of the data quality criteria and the objectives of each activity and stage of the program withing the framework of the available resources, and use less expensive alternatives if appropriate.

b. Considering alternatives to TO or IO methods for specific compounds for which these methods are known to be inadequate(e.g., DNPH-based methods for acrolein or SUMMA canisters for polar compounds).

c. Developing a clear set of criteria that will guide the selection of alternate methods in the future as sampling/analytical methodology evolves.

3.1.3 Question 3. The Monitor-to-Model Evaluation Protocol

3.1.3.1 Question 3(a)

Do the data analysis approaches provide enough information to allow appropriate interpretations of model results to support the development of model improvements in the future and to assist with the design of the national monitoring network?

The Subcommittee generally agrees with the proposed data analysis approach as an initial step. However, the Subcommittee notes that any effort at model validation will be limited by a number of factors, including lack of multimedia measurements for relevant air toxics, the limitation of the analytical methods for measuring certain HAPS, and the well-recognized limitations of emission inventories. The Subcommittee is very concerned about the exclusion of models for multi-medial pollutants that move and accumulate in soil, water, and the food chain, and the consequence of this exclusion for achieving the objectives of the program. The Subcommittee recommends that the Agency:

a. develop specific performance assessment criteria for the comparison of model estimates and measurements.

b. perform Monte Carlo simulation of model outputs as a means of parametrizing the model estimates for future comparison with measurements.

c. prioritize the data analysis approaches from the least-sophisticated but more easily understood and conveyed (e.g., rank correlation) to the more sophisticated but not as easy to explain (e.g., tests of medians and quartiles).

d. Further develop the stratification approach to data analysis to include source categories, type of source, type of terrain and meteorological variables.

e. Develop a set of metrics to evaluate in detail the potential causes for disagreement between model estimates and measurements.

f. Consider that persistent HAPS may have different dispersive characteristics

g. Start to develop support for inclusion of multi-media monitoring and modeling efforts.

h. Consider use of additional statistical analysis methods such as multivariate regression and test of medians and quartiles.

3.1.3.2 Question 3(b)

Are there some HAPs for which these approaches appear inadequate? If so, can the Subcommittee suggest alternative approaches for these?

The Subcommittee noted that there are HAPS for which the Model-to-Monitor comparison approaches will be inadequate. The Subcommittee expressed particular concern that multimedia pollutants are excluded from both the monitoring and modeling framework. In addition only inhalation is considered as an exposure route. The Subcommittee recognizes that there are not sufficient resources to include multimedia pollutants in the first phase of the monitoring efforts. Providing adequate attention to multimedia HAPs will require a multimedia monitoring strategy and multimedia exposure models.

In addition to the concern expressed about the inadequacy of the approach for multimedia pollutants, the Subcommittee noted that the Model-to-Monitor approach could have inadequacies for the following HAPs categories for the reasons noted:

a. Those for which problems exist with current sampling and analytical methodology (e.g., acrolein, acrylonitrile).

b. Those for which adequate detection levels are attainable but for which current analytical methods are significantly above the exposure-level of concern.

c. HAPS having uncertain, or poorly established, emission inventories.

d. Because multi-media pollutants can be locally cycled, the will be observed by monitors but their real concentrations will not be captured in local emissions inventories.

3.1.3.3 Question 3(c)

As noted in the paper, annual-average concentrations and comparisons to modeled estimates can be uncertain when a large percentage of the measurements are below the method detection limit (MDL). To estimate annual-average concentrations from monitoring data, EPA generally substitutes half the MDL. Does the Subcommittee suggest any alternative statistical approaches?

The Subcommittee believes that substitution by one-half the MDL, while a fairly robust approach, may not be appropriate for all air toxics and all situations. The Subcommittee recommends that:

a. The Agency develop a set of criteria for what would constitute an acceptable fraction of below detection concentrations as part of the data quality objectives, sampling and analytical methods, and spatial and temporal sampling considerations.

b. If possible, incorporate the need for obtaining larger proportion of above detection values as a criterion for the sampling/analysis approach.

c. Select the MDL substitution method that best fits the amount and distribution of the data, the fraction of values that are below the MDL, and the specific objective for which the data will be used.

The Subcommittee recommended that laboratories report all data with the associated uncertainties rather than an MDL, because the MDL is a variable in and of itself. This way, end users of the data can decide for themselves the appropriate uncertainties for utilizing the data. For data sets which already contain MDL values, the Subcommittee makes no specific recommendation, but offers several approaches for statistical treatment of MDLs.

3.2 Detailed Responses to Question 1

Does the air toxics monitoring concept paper describe appropriate air toxic monitoring objectives and principles, particularly ones that will permit the collection of monitoring data to support the initial National Air Toxics Assessment activities?

The objectives of the National Air Toxics Assessment (NATA) program are to "help EPA identify areas of concern, characterize human health and ecosystem risks and track progress." The activities of the NATA program are reiterated here for the purpose of discussion here. The specific aims are:

a. measurement of air toxics emission rates from individual pollution sources,

b. compilation of comprehensive air toxics emission inventories for local, state, and national domains,

c. measurement of ambient concentrations of air toxics at monitoring sites throughout the nation,

d. analysis of patterns and trends in ambient air toxics measurements,

e. estimation of ambient and multimedia air toxics concentrations from emission inventories using dispersion and deposition modeling.

f. the estimation of human and environmental exposures to air toxics, and

g. the assessment of human and environmental risks due to air toxics.

Specific aims c-g should be considered as tactical approaches for achieving aims a and b, the overall primary objective of NATA.

Among the activities that derive from these specific aims, the concept paper indicates that ambient monitoring data are needed to assess the air toxics inventory, air toxics modeling and trends in HAP concentrations. The Subcommittee recommends that these uses of the monitoring network be prioritized to provide a sound rationale for the study design. Establishing a well-designed ambient network(s) for estimating ambient concentrations could also assist in determining exposures across the nation when combined with appropriate models. However, from what was presented at the meeting, it is not clear how the EPA will know:

Whether the ambient network is well designed?

What criteria would be used to assure that the network is adequate?

These questions could be better answered if priorities for the network are clearly defined.

The draft Air Toxics Monitoring Concept Paper identified a number of important scientific issues. However, a well thought-out, step-wise process that addresses these issues and focuses on achieving the primary monitoring objectives should be developed. The EPA could begin this process by defining as a goal the acceptable levels of uncertainties and then the acceptable data quality that is needed for achieving the goal. These uncertainties should be communicated in all data analysis activities or assessments based on these data.

If monitoring data are to be used to estimate exposures, measurement error is of particular concern because it can be a potential source of bias. Exposure measurement error can occur when using ambient air measurements to estimate human and ecological exposures. Because of these errors, data provided by centrally located monitors rather than exposures measured on individuals could affect the relative risk estimates in ways that are difficult to predict. The possibility of such errors causes uncertainty about the true magnitude of the estimated effects of individual air toxics on health. For a more complete discussion of exposure measurement and its relationship to sources, see APPENDIX A: Exposure Measurement Issues

The proposed air toxics monitoring program is based on the assumption that state and local agencies will make most of the measurements, and that EPA will provide some funding to the States for this. Although state and local agencies have the expertise to carry out field sampling programs, with adequate guidance from EPA, other tasks, such as personal monitoring, methods development, and other research-oriented objectives, are not typically within the ability and authority of the State and local agencies.

3.2.1 Question 1(a)

Does the Subcommittee believe that we identified the most important uses for ambient air toxics data? Are there other types of air toxics monitoring data or data uses that should also be identified for near-term or for future air toxics monitoring activities?

To address this question the Subcommittee considered current, near-term and future uses of air-toxics data. These uses were also evaluated has being primary, that is directly related to the specific aims of NATA, and secondary, of use to other EPA and state programs.

3.2.1.1 Current Primary Uses for Air Toxics Data EPA has identified many important uses of the data, including: model parameterization, model evaluation, trends analysis for GPRA, background measurements, source characterization, national air toxics assessments, and residual risk program assessments.

Because the National Ambient Monitoring Network will be the primary source of air toxics data for many uses, the data should be collected in a manner appropriate for multiple uses. The Agency's Quality System and the SAB's reports on secondary uses of data may be helpful in determining how this may best be done.

Some additional uses of the data will naturally evolve from the 1-2 year pilot program.

From a regulatory perspective, data gathered during this pilot phase will:

a. characterize ambient concentrations and deposition in representative monitoring areas,

b. provide a "reality check" to dispersion and deposition models, and

c. decide on the appropriate quantity and quality of measurements in a national monitoring network.

The Subcommittee agrees that multiple sites operating over at least a one year period in several different regions of the country will be needed to adequately characterize a given monitoring area and provide a minimal "reality check" on current models.

The Subcommittee agrees that EPA should focus on the "Urban HAP List" developed as part of the Integrated Urban Air Toxics Strategy. These chemicals are of high priority both for health risk and because of frequency of detection. EPA may also wish to consider groups of substances that serve as "fingerprints" for specific source emissions because this will make the initial data sets amenable to alternative source-receptor modeling as an additional "reality check" on the dispersion and deposition models.

Although the Subcommittee members expressed a range of views on the selection of HAPs to be measured during the pilot phase, there was consensus that the target list should reflect the following issues:

a. practicality of measurement,

b. relative toxicity and ambient concentrations of the compound, and

c. inclusion of more compounds as the monitoring program matures.

Some thought the initial measurement phase should include as many compounds as is practical, given method limitations and probabilities of detection, and that this list should include at least the 18 core compounds identified in the concept paper, with possible addition of other multi-media pollutants that include an air pathway. Others advocated that the initial measurement phase focus on fewer compounds, selecting those whose behavior is most consistent with the assumptions of the dispersion and deposition models to be tested.

3.2.1.2 Potential Secondary Uses of Air Toxics Data Given the scarcity of air toxics monitoring data, EPA should be prepared for the likelihood that the publicly available data from this project may be used for additional analyses. The Subcommittee notes that some of these uses could contradict Agency policy for the use of monitoring data as specified in 40 CFR Part 51 of April 21, 2000 section 10.2.2. Furrthermore, there is the potential for inappropriate uses of presently available dispersion models which cannot confidently distinguish contributions from individual sources to ambient monitors in most cases. By providing sustained attention to the Agency's quality system, available modeling techniques and documentation, EPA can provide a scientific basis for determining whether or not the data can be used appropriately to support secondary uses. The Subcommittee expects EPA will restrict itself to appropriate uses and hopes that others will do the same. However, some individuals and organizations may be less sensitive to the nuances of appropriate secondary uses of data. If data are used inappropriately to reach conclusions that are not scientifically defensible, then EPA will find its sustained and documented attention to the Agency's quality system provides a scientifically credible defense against faulty analyses. The following are areas where the Subcommittee anticipates secondary uses of the data from this study:

a. Identifying un-permitted emissions not found on current inventories, such as fugitive air emissions, air emissions from commercial underground injection sites, etc.

b. Determining potential impact of siting new facilities using chemicals monitored by the project,

c. Determining if reported large emission reductions have actually occurred, or if reported reductions were due to changes in calculation method,

d. Supporting and assessing the impacts of other EPA programs, such as OSWER's voluntary reduction program for persistent bioaccumulative toxics (PBTs),

e. Supporting the fish advisory program's efforts to determine whether there are advisories for certain chemicals in areas where air concentrations indicate that there may be problems.

3.2.1.3 Other near-term and future uses for data In planning their monitoring program, EPA should make use of existing data on air toxics, but be aware that data quality criteria be assigned based on measurement quality and the time frame. In the initial pilot phase of the monitoring activities and as an aid in the selection of sites nationwide, data from EPA-funded and non-EPA funded research studies could also be used even if it was not collected according to EPA-approved monitoring methods.

If resources are available, monitoring of multiple media for chemical contaminants (e.g, PCBs, mercury, dioxin) can strengthen the exposure models which are the basis for risk assessments required by the Clean Air Act (e.g., residual risk assessments). These media should include: soil, surface water, sediment, fish and plant foliage.

Data from this study can contribute to the assessment of models other than ASPEN, including TRIM.FaTE and hazardous waste combustion exposure models. These data could also be of use to those conducting personal exposure studies. EPA collaboration with those organizations planning large National Institute of Health (NIH) type health surveys could be of benefit to both organizations.

3.2.2 Question 1(b)

Does the Subcommittee believe that neighborhood-scale monitoring is appropriate for evaluating ASPEN air quality predictions and later for developing long term ambient air quality trends? Are there other appropriate monitoring scales, perhaps for other data uses, that the Subcommittee would suggest?

The Subcommittee agrees that neighborhood scale monitoring is the correct choice for measurements in urban areas during the pilot stage of this program. Neighborhood-scale monitoring is meant to be representative of a 0.5 to 4 km horizontal scale. This is appropriate for comparison with ASPEN results for urban census tracts, which average 2.3

km^2 in area, according to the Cumulative Exposure Project study. Neighborhood-scale monitoring is also a good starting point for tracking long term ambient air quality trends in urban areas. However, over the longer term, the monitoring scale used for toxic air pollutants should be guided by both the objectives of the monitoring program and the characteristics of the pollutants being monitored, e.g. their reaction or removal rates and the distribution of sources that determine the spatial variability in their concentrations. Different scales may be used depending on the compound class and purpose.

In many cases, temporal variability in concentration will be more important that spatial variability (for example, a background site or a city with few sources), and location could be specified at the urban rather than neighborhood scale. However, for some comparisons, the neighborhood scale could be too large because of the impact of significant sources in very-close-in areas. In these cases, source-receptor characteristics will dominate siting decisions. In these situations, EPA may wish to take a multi-tier approach to site selection, based on careful evaluation of potential site conditions and the multiple uses of the data.

Conducting limited microscale monitoring to assess spatial variations within a census tract is not essential during the pilot stage of the program. Experiences from the recent MATES-II study in southern California have shown that the use of microscale monitors to detect areas with localized higher concentrations of toxic air pollutants is more difficult than usually perceived. Distances in the order of 200 meters or so can have concentration gradients of up to two orders of magnitude. Thus an array of monitors may be needed to depict influences from point sources. Therefore, the Subcommittee recommends avoiding siphoning off resources for microscale monitors and placing the initial focus of the national monitoring program clearly on the neighborhood scale.

3.2.2.1 The Problem With Rural Census Tracts In contrast to urban census tracts, the average area of the rural census tracts considered in the Cumulative Exposure Project was 242 km^2. Therefore, background monitors representative of larger areas may be needed for comparison with model results for rural census tracts. Because OAQPS plans to include "rural" monitors in the pilot stage, the revised concept paper should discuss guidelines for siting them.

Neither the monitoring scales considered in the concept paper nor the census tract divisions used in ASPEN modeling are appropriate for assessing ecological exposures. Therefore, in the future, the program may wish to consider establishing some monitors for

ecological exposures, at which time siting criteria that are appropriate for this purpose should be developed.

3.2.2.2 Co-Location With Other Monitors EPA should be cautious about recommending the co-location of air-toxics monitors with existing PAMs sites because these locations may not be optimal for assessing air toxics exposures.

3.1.3 Question 1(c)

Does the Subcommittee believe that a basic 24-hour sample taken at a frequency sufficient to fulfill the objectives of the program is adequate to provide the model reality check and supply data for the characterization of ambient hazardous air pollutant (HAPs) concentrations? In particular, what are the Subcommittee's thoughts on the use of 24-hour samples collected once in 12 days for model evaluation and at a more frequent (say, 1-in-6 or 1-in-3 day) schedule in the future for trends assessment at permanently located monitoring sites?

The Subcommittee believes that the 24-hour sampling frame, while reasonable in principle, may not always be consistent with the objectives of NATA for all compounds and foreseen uses of the data.

In some cases a longer sampling period may be more useful. For example, longer duration samples may help elucidate the relationship of long-term exposures to chronic health effects. Therefore, sampling for 24-hour periods may not be appropriate if the purpose is to compare the mean of the measurements with a model estimate of average concentration for a longer period, such as one year. Assuming that there are no sampling duration-related artifacts, and that the monitoring approach permits it, a longer sampling time (e.g., 1 week) is beneficial. The longer sampling time would result in a smaller fraction of the data being below the MDL, lower cost, and the quasi-continuous samples will provide even better comparison data.

Sometimes a shorter sampling period will be more appropriate. For example shorter duration samples may help elucidate the relationship of brief exposures to acute health effects. If the measurements will be compared with a model estimate of the maximum 24-hour concentration over the period of a year then shorter-term sampling is a better approach. Sampling for 24-hour periods may also be appropriate if future trend analyses address concentrations at the upper end of the distribution. The highest

concentrations, which characterize the upper tail, occur infrequently. As a result many measurements are needed to characterize the tail.

Because chemistry affects how compounds behave in the environment, what effects they cause, and how they can best be measured, sampling duration could be considered compound specific, or perhaps compound-class specific. For example, for sorbent-based methods, but not canister sampling, sampling over longer time periods will reduce the detection levels because a larger quantity of sample is collected. To address this issue, the Agency must consider the specific uses of the data as foreseen now, brainstorm to determine potential future uses, and then adopt sampling times that satisfy the requirements of all, within the available resources.

EPA's *Guidance for the Data Quality Objectives Process* (QA/G-4) should be consulted as a guide to planning for projects where the objective of the study is to collect environmental data in support of an Agency program. This is important where the results of the study will be used to make a specific decision. Data Quality Objectives (see Appendix B) developed with the above considerations can help guide decisions regarding:

a. the appropriateness of taking 24-hr samples,

b. the appropriate frequency to use for sample collection (every 3, 6, 12, etc. days),

c. the number of monitoring stations needed to produce sufficient data to reduce uncertainties to an acceptable level, and

d. the spatial distribution of monitoring stations needed with respect to the population and environment. In the absence of a rigorous approach, it is likely that the modeled results will have large unacceptable and ill-defined uncertainties.

3.1.3.1 Time Resolution of Samples The precision/bias desired for the models and for risk estimates should also be a factor in setting the frequency of sample collection. Too long a cycle may miss episodic emissions and very short sampling cycles will require more resources.

Consider the case of a power plant as an example. During the summer, electricity demand can vary as much as 40 percent from day to day, especially during heat waves. To meet that power demand, electric power companies increase generation at their plants, many of which burn coal or fuel oil, and these plants emit VOCs in addition to nitrogen and sulfur oxides, mercury, and other metals. VOC emissions are a function of the amount of fuel consumed, which increases during increased demand. Unfortunately, although nitrogen oxide emissions also increase, these emissions are more a factor of the burn

temperature, air flow for combustion, and duration of burn, and so are not a good proxy for VOC emissions (most utility-owned power plants continuously monitor their nitrogen oxide emissions). Therefore, with short-term heat waves, the one-in-twelve monitoring proposed will likely not be frequent enough to capture the short-term large increases in VOC emissions. According to EPA's AIRS database, electric power plants burning coal and oil are among the top producers of VOCs nationally, and will likely constitute the major source of VOCs in some neighborhoods or rural areas. (National Environmental Trust, May 1997)

3.1.3.2 Seasonal Variation and Annual Average Currently available data from six-day sampling cycles could be used to determine the effect on precision of less frequent sampling. This could be done by examining the impact on the autocorrelations for each HAP. For HAPs that vary by season a sampling frequency greater than 7 samples per season (1 in 12 days) may be needed in order to precisely describe variability and to obtain an accurate annual average.

Diurnal variation and resulting health effects are also issues for certain HAPS. Where these issues are important, EPA could consider selecting a number of sites (e.g. 1-3 sites) for 12-hr day/night samples. Although the Subcommittee does not favor co-location of monitors (see response to question 1(b) above), the Subcommittee suggests that PAMS data should be analyzed for diurnal information regarding the toxics compounds measured in that program. In some situations, sampling done on a 1 in 12 basis may preserve resources and allow a larger number of sites to be monitored. More frequent sampling at a selected number of sites (e.g., 1-3) should have 1 in 6 sampling; and perhaps one site should have 1 in 3 sampling to help elucidate the potential importance of variations. Future long-term monitoring should be based on results of the pilot study.

3.3 Detailed Responses to Question 2

Does the air toxics monitoring concept paper present a reasonable phased strategy to design a national air toxics network?

3.3.1 Question 2(a)

Does the Subcommittee believe that data analyses of existing air toxics monitoring data, coupled with focused pilot studies for a core set of HAPs

in a limited number of areas are appropriate first steps in the design of the national network. What additional or alternative approaches are suggested?

In general, the "data analyses of existing air toxics monitoring data, coupled with focused pilot studies for a core set of HAPS in a limited number of areas are appropriate first steps in the design of the national network." However, these analyses should not be limited to the generation of statistical summaries. They must also focus on how the dimensions of existing measurements in space and time relate to the dimensions of the decision units of interest.

Any statistical analysis must not only supply a data summarization, but also supply a description of how these summary statistics are related to the entity about which inferences are desired. This entity is the “Decision Unit”. It is critical to clearly define the spatial and temporal boundaries of this entity before sampling and/or ascribing any meaning to pre-existing sampling data. This is step 4 of the DQO process. (See Appendix B for a description of the DQO process.)

Conversely, when employing existing monitors, a clear description of the spatial and temporal boundaries of the entity (air volume) actually monitored is essential. If these boundaries do not coincide with those of the Decision Unit, then the data generated by the monitor are of little value in making inferences regarding the Decision Unit.

3.3.1.1 Data Gaps and Pilot Studies Such analyses will be quite helpful in identifying data gaps and planning for focused pilot studies. EPA and the Subcommittee have already identified a few data gaps that may merit pilot studies. For example, a pilot study that would address vertical variability may be helpful to determine if monitors are adequately representing exposure critical for ecological studies as well as human health risk assessment.

3.3.1.2 Mobile and Stationary Sources The air toxics monitoring concept paper presents a reasonably phased strategy for designing a national air toxics network and the proposed pilot study is a good step toward designing this. In choosing sampling sites for this network, EPA should consider selecting sites that will be representative of emissions from mobile sources, sites that will be representative of emissions from stationary sources, and sites where there are contaminants from both mobile and stationary sources.

3.3.1.3 Personal Monitors Personal monitors do not have the sensitivity that is currently achieved by ambient air sampling. In addition, personal monitors are not likely

to be used on a large scale. However, characterizing exposure requires both indoor and outdoor concentration levels. Therefore, EPA should consider developing outdoor/indoor penetration factors for selected air toxics in parallel research programs, perhaps those associated with PM2.5 studies.

3.3.1.4 Conveying Uncertainty It is very important to convey the level of uncertainty. This includes uncertainty in the models and the risk assessment, as well as sampling/analysis. In this way, reasonable expectations of precision can be communicated to the public.

3.3.2 Question 2(b)

Given that the State and local agencies have been measuring ambient air toxics with the Toxic Organic (TO) and Inorganic (IO) methods as described in the paper, are these methods appropriate for the continued routine monitoring of the target Urban Air Toxics Strategy compounds in a national monitoring network? If not, are there alternative methods which the Subcommittee would recommend?

The technical criteria must be compatible with the data quality objectives of the study and EPA's *Guidance for the Data Quality Objectives Process* (QA/G-4) is a useful approach to working these issues out.

The methods proposed for the network are the currently accepted “gold standard” procedures tested and approved by the EPA. For example, if the purpose of the study is to understand exposures near the current MDL, these methods are probably necessary. This does not mean that using these methods is always the best choice; other uses of the data may have objectives more compatible with other methods. To make an analogy to medical testing, screening tests with relatively high false positives are often used in situations where large numbers of measurements need to be made; then the more expensive confirmatory tests are performed where the screening tests suggested there might be a problem.

The DQO process enables the Agency to think through its objectives and the costs of meeting them in a systematic way. Studies that involve collection and analysis of large numbers of samples often need to balance factors such as the number of samples that can be obtained at any given site, and the number of sampling locations and test methods. Because the development of a sampling network involves a series of decisions, the

balance among these considerations may change from decision to decision. Sometimes resources are inadequate to meet the original goals. If so, decisions must be made about altering the goals, the approach or the resources.

3.3.2.1 Matching the Method to the Need In this section, the Subcommittee is addressing the question, "which are the best methods at each stage of the project." The idea is to suit the test methods to the budget, the intended use of the data at each stage and how it may impact the next stage of the project.

Site Selection. Where the site selection is the issue, a screening sampling method may be the most appropriate approach because it is more important that the sampling locations are identified than that the individual measurements are exact. The most rigorous methods could be used to analyze samples from a few selected sites to provide a measure of reliability for the screening tools.

Accuracy Once the sites are selected, then the accuracy of the measurement may become the driver for the sampling and analytical approach. If so, then the most rigorous method is the more desirable approach.

Once the sites to be included in the network are selected, then "gold standard" methods could be applied routinely. In some cases, improvements could be made to the TO-sampling methods. For example, silica-lined Summa canisters may provide for more stability of polar species and there is evidence that they are suitable for target HAPS such as ethylene oxide and acrylonitrile.

While the current TO methods are generally appropriate, the utility of silica lined canisters should be explored in order to extend the TO-14,15 approach to polar compounds (ethylene oxide, acrolein, acrylonitrile) that are difficult to sample with current stainless steel canisters. Frequent performance audit checks of laboratories doing TO-methods should be performed to insure that these laboratories are meeting accuracy, precision and detection level criteria.

Methods and Costs Influence the Number of Sampling Sites Based on these considerations, the Subcommittee recommends that the Agency reconsider type of possible sampling methods. The Subcommittee notes two examples. One, if the Agency compares the cost of integrated vs. real time monitoring methods , and finds real time methods to be satisfactory as screening tools and less expensive overall than integrated sampling, then they may be a more suitable choice for the initial phase of selecting

sampling locations. Two, passive sampling may be appropriate for some HAPS over longer sampling periods.

3.2.2.2 Selecting the Appropriate Method The Subcommittee suggests EPA use the following key criteria to assess the appropriateness of methods for use in monitoring the target Urban Air Toxics. These criteria include the ease of sample collection by minimally trained individuals, the ease of sample transport and storage, the stability of the pollutants in the collected state, the detection limits of the method (i.e., the combination of sample size collected and the sensitivity of the instrumental measurement technique), the method precision and bias, general ruggedness of the method, the existence of other extant data collected with the same methods (comparability issues) and cost of analysis per sample.

Advantages of Current TO- Methods The TO-4A, -13A, -14A and -15 methods for collecting pesticides/PCBs, BaP/PAHs, and VOCs (non-polar and oxygenates) are well established procedures. Most state and local agencies have acclimated to using these procedures and are experienced with these methods. A considerable body of data now exists with these methods and the precision and bias are well understood. It is also likely that after the DQOs are defined as described above, the precision and bias will be acceptable for the objectives of NATA. Sample transport and storage and the stability of the pollutants in the collected state will also likely meet the criteria for this program. The use of GC/MS for measuring organics is highly recommended since the instrumental method yields data with the fewest false positive results, i.e., misidentification is very low. It is however, strongly recommended that modern GC/MS instruments be employed, since their sensitivity is as much as 50-fold higher than those 5 or more years old. This will help reduce the percentage of non-measurable values for each of the HAPs, a concern in the use of modeling techniques for estimating exposure. GC/MS employing full scanning approach should be used which will permit the very valuable opportunity of retrospectively examining data for other or new HAPs in the future. While a major goal of NATA is the ability to perform a trends assessment for the target list of HAPs, full scan GC/MS acquisition of data will also permit identifying emerging pollution problems.

Disadvantages of Current TO- Methods The preparation of collection materials and the collection procedures prescribed in these methods are labor intensive. The sample work-up for the pesticides/PCBs and BaP/PAHs is also labor intensive.

Taking into account the entire effort to implement TO-4A, -9A and -13A methods, it is not surprising that the costs are high per sample. Because the VOCs are measured directly from the collection device, i.e., no work-up is needed, the cost is somewhat less. Never-the-less, the cost for analyzing thousands of samples collected throughout the monitoring network will be very expensive.

Comments on Other Methods TO-11A for the collection and measurement of formaldehyde and other aldehydes is also a well established method. The method employs high performance liquid chromatography with UV detection. The drawback of the method is that it is limited to a target list of aldehydes. Since modern HPLC/MS technology provides sensitive measurements, It is recommended that HPLC/MS be used for measuring the dinitrophenylhydrazine derivatives of aldehydes. This approach will permit the opportunity to identify emerging trends in new pollutants that would be missed with the current method of detection.

IO-3 is used for collecting fine and total suspended particulate matter on filters and for the analysis of metals. Similarly to its counterpart TO methods, this method is well defined and used method. Alternatives to X-ray fluorescence measurement of metals do exist; however, many of them require sample digestion (destruction).

General Comments The Agency should also encourage the use of state-of-the-art instrumentation with each of the analytical methods used in NATA. For example, modern mass spectrometers have improved sensitivity such that a factor of over 50 can be achieved compared to instruments manufactured in the mid-1990s. This added sensitivity will permit the measurement of HAPs at lower levels than in past years yielding fewer non-detects. Also high throughput systems will be available in a few years that will permit the simultaneous analysis of 3 to 6 samples, thus lowering the analysis costs significantly. The Agency should encourage analytical laboratories to stay abreast with these developments for use in NATA.

3.4 Detailed Responses to Question 3 on Model-to-Monitor Comparison

3.4.1 Question 3(a)

Do the data analysis approaches provide enough information to allow appropriate interpretations of model results to support the development of model improvements in the future and to assist with the design of the national monitoring network?

The data analysis approaches provide a good starting point to evaluate the performance of the ASPEN model and to facilitate improvements to the monitoring network. The data analysis approaches provide the opportunity to calibrate the ASPEN or other models with monitoring data. Once calibrated, ASPEN could be applied with greater confidence in situations where the monitoring data are limited. Similarly, calibration will allow EPA to use the ASPEN (or other models) to design monitoring networks.

To define performance objectives for models used to predict air toxic levels in time and space, EPA must first establish acceptable uncertainties. There are two broad categories of uncertainties: those that are associated with the model's ability to describe the system of interest and those that are associated with the input parameters to the model. In general, model uncertainties are larger and more difficult to assess than input parameter uncertainties. Both should be evaluated to the extent possible in order to gain confidence in a model's performance for use in this program. This will require understanding which input parameters are the most influential in determining the outcome or, if not known, performing sensitivity testing of the models. The potential sources of uncertainties in ASPEN have been identified; however, the sensitivity of input parameters leading to the magnitude of uncertainties is unclear. Once this knowledge is in-hand, effort should be devoted to obtaining high quality data for the more significant input parameters consistent with achieving the desired level of confidence. The results from modeling efforts should display confidence intervals.

3.4.1.1 Proposed Data Analysis Methods

Assessment Tools The development of performance assessment methods is key to evaluating the model relative to the monitoring data. In addition to straight-scatter plots and box plots, EPA used three performance assessment tools in the report and at the meeting--probability plots, Spearman's rank correlation test of measured and modeled concentrations, and a point-to-range test of monitored concentrations with the model estimates for the county in which the monitor is located. The first two tools are for classification analysis, whereas the latter makes more explicit use of the exact concentrations. The probability plot is based on the transformation of model measurement pairs into a hit (1) or miss (0) score with all location/measurement pairs plotted on a logistic curve. The rank correlation compares the ranks of the monitored averages with the ranks of the model estimates, to see if the model and monitors rank the sites similarly. The point-to-range test compares the point observation to the range of modeled concentrations for the selected county.

Priorities All three methods have clear capabilities and limitations. The Subcommittee found the rank correlation approach easier to carry out and to communicate to decision makers and thus suggested it should be given highest priority as an evaluation tool. The Subcommittee suggested that the rank scatter-plot would be a useful adjunct for assessing the rank correlation. Scatter plots of absolute concentrations are also useful; these could be done with model estimates for neighboring census tracts as well as for the exact monitor location.

Second priority should be given to the point-to-range test. The point-to-range test is useful because it is an "on-the-ground" test, and the Subcommittee recommends its use with one caveat. Because of the variation in census tract size, the observation of a miss in one range may not be the same as another. This might cause confusion for those trying to interpret the results. It may be more useful to compare the monitoring average to the range of model estimates for several neighboring census tracts than for the range for the county that includes the monitor, given the variability in county size and monitor location within a county. The monitor may be located near the boundary of a census tract and/or county.

The use of probability plots offers the potential for more statistical sophistication, but had the disadvantages that the method is difficult to explain and that it is difficult to interpret the results. This is a significant disadvantage for information that must be communicated to a broad range of decision makers.

3.4.1.2 Use of Stratification Stratification to evaluate monitoring data is very useful and should be strongly encouraged. The stratification of samples is a reasonable way to discern causes of model and monitoring mismatches, with the caveat (which EPA recognizes) that stratification "slices a thin pie [of samples] even thinner." The currently proposed stratification variables are reasonable, these are:

a. urban vs nonurban,
b. geographic/climatological region
c. pollutant level, and
d. source-oriented monitors vs others.

However, the Subcommittee suggests that additional factors be considered in developing the stratification, such as:

a. wind speed,
b. terrain,
c. season, and
d. source categories (such as point, mobile, area, etc) .

The terrain would be difficult to represent as a single variable for a census tract; although a digital elevation model may be helpful, it is also a great deal of work. Subcommittee members vary on their views of which stratification variables suggested by EPA are the most important. Some favor pollutant level (above the MDL) and geographic region, while others think geographic region as it relates to meteorological factors could be informative, and stratifying by pollutant concentrations could give significant hints about whether the model or the measurements are in error. Errors in the emissions inventory may overwhelm any differences in model output due to the presence of terrain features or other stratification variables, but this assumption would need to be tested.

The stratification provides the opportunity to bin observations in a way that allows the monitoring and model results to be linked to similar attributes. For example by evaluating monitor/model concentration ratios among all sites with low (or high) median wind speeds we can look for any systematic problems in the model performance under the low-wind conditions.

3.4.1.3 Model Diagnostics and Model Reliability Although the stratified model comparisons will give a flavor for how well the models work relative to monitored concentrations, these comparisons will not provide detailed diagnostics on model performance. Thus, the comparisons can not be expected to provide enough diagnostic information to determine how the ASPEN or other EPA models may be corrected or improved. The Subcommittee suggested ways to address this issue.

First, EPA should identify the purpose for the model and the expected reliability of that performance. The conceptual plan for the monitor-to-model evaluation does not yet include quantitative metrics or criteria to identify the differences in measured and modeled average concentrations that would be a cause for concern. Such a metric could be used to indicate the need for a more detailed evaluation of the model structure. To start examining problems that might arise from meteorological inputs or treatment of dispersion in the model, it will be necessary to investigate model-monitor comparisons at finer time scales than the proposed annual average comparisons. One way to address this is to work with a subset of chemicals that can be linked to the same source, such as benzene, toluene, and ethylbenzene from fuels and see how they agree in the monitoring, model and source data. One could look at comparisons between monitored and modeled ratios of pairs of contaminants at any time scale -- finer time resolution isn't necessary. If ratios don't match, one might suspect the emissions inventory, since meteorology should affect pairs of pollutants from similar sources in similar ways and thus cancel out of the ratio.

At concentrations at or below the MDL, the ASPEN model may not be easily evaluated when there is a large percentage of non-detects in the monitoring data. This is unfortunate, given that detection levels for some chemicals, such as some VOCs, are several orders of magnitude higher than the critical health values.

Measured concentrations and modeled estimates may be expected to be different if the pollutant is a persistent and is distributed into other media. In this case monitors may reflect long term emissions from soil or vegetation which are not predicted by ASPEN. The nine HAPs proposed for the Model-to-Monitor comparison should include a persistent, bioaccumulative organic pollutant such as a 3-4 ring PAH or PCBs unless it is demonstrated that the dispersive characteristics of these chemicals are similar to at least one of the nine PAHs.

3.4.1.4 Alternate Methods for Comparing Models to Measurements

The Subcommittee offered suggestions for alternate methods for evaluating the congruency of predictions and observations.

a. Multivariate regression can be applied similarly to both the monitoring data and the model predictions. That is, premises can be tested for dependence of observed or predicted concentrations on variations in source, climate conditions, terrain factors, source category, etc.

b. Gilbert and Simpson (1990) have proposed a method for evaluating soil concentrations, which has been adopted by the USEPA (1994) and the Nuclear Regulatory Commission (1995). This approach involves a test of medians and a quantile test on extremes.

Both methods require a “statistical” distribution of observed annual average concentrations and model predictions. These entities result from temporal and spatial series with perhaps the superposition of measurement error. Therefore, it may be difficult to define what we have when we say we have congruency within some geographical area.

c. The Canadian NTRI data are more detailed than NTI data with respect to contaminant emissions, such as PERC, PAHs, and dioxin, from some sources, and these could facilitate the identification of gaps in the NTI.

3.4.2 Question 3(b)

Are there some HAPs for which these approaches appear inadequate? If so, can the Subcommittee suggest alternative approaches for these?

Problems exist for HAPS in the following categories:

a. Those for which problems exist with current sampling and analytical methodology (acrolein, acrylonitrile). These may be resolved by considering alternative techniques such as substitution of silica-lined canisters for stainless steel canisters in TO-14, TO-15.

b. Those for which detection levels are attainable but current analytical methods are significantly above the 10^{-6} risk based concentration. This is particularly true for the class of volatile organic compounds. Since there is not any likelihood that the method detection limits will be significantly reduced in the next few years the issue of dealing with "non detects" becomes critical here. Some form of sensitivity analysis should be carried out on the effect of "non-detects" upon the annual average concentrations derived from monitoring data.

c. HAPS having uncertain, or poorly established, emission inventories. Values should be assigned to designate the quality of the emission inventories for specific HAPS and their contribution to the uncertainty of modeled concentrations.

Problems may exist for compounds that can be adsorbed on soil, biomass, or water surface and then locally cycled and re-emitted in such a manner that they are not included in local emissions inventories. This categories would include semi-volatiles, organic species such as polychlorinated byphenyls and polycyclic organic matter.

The Subcommittee remains concerned that multimedia pollutants are excluded from both the monitoring and modeling framework. In addition, only inhalation is considered as an exposure route. This excludes several classes of HAPs -- semivolatile compounds that are transferred through food chains. The Subcommittee recognizes that there are not sufficient resources to include multimedia pollutants in the first phase of the monitoring efforts. However, the Subcommittee is concerned that, because they are left out of the first phase, they will not be considered in the future and absence of information could be interpreted as the absence of a problem. Providing adequate attention to multi-media HAPs will require a multimedia monitoring strategy and a multimedia exposure model.

3.4.3 Question 3(c)

As noted in the paper, annual-average concentrations and comparisons to modeled estimates can be uncertain when a large percentage of the

measurements are below the method detection limit (MDL). To estimate annual-average concentrations from monitoring data, EPA generally substitutes half the MDL. Does the Subcommittee suggest any alternative statistical approaches?

The Subcommittee observed that the substitution by half the MDL, which is a fairly robust approach and commonly used, may not be appropriate for all air toxics and all situations. In reviewing this issue the Subcommittee noted that dealing with data below the MDL is an important issue and must be confronted. However, the subcommittee offers cautions about interpreting the MDL and how this impacts methods for dealing with monitoring data below the MDL.

3.4.3.1 The Importance of Dealing With Measurements Below the MDL

At and below the MDL, concentrations cannot be measured reliably and this makes certain analyses difficult. The proportion of non-detects and how they are treated in the analysis can change the estimate of the averages. Where toxic potency factors are high, even small changes in the estimate of the averages can result in substantial variability (and additional uncertainty) in the calculated risk. Where the proportion of non-detects is large, the ability to compare modeled versus measured concentrations is impaired because the measures of central tendency and the variability depend upon the estimates of distribution parameters. As a result, these estimates become more and more uncertain as the fraction of the concentration data that is below the MDL increases.

3.4.3.2 Cautions The appropriate interpretation and use of the MDL is not a simple matter.

a. Because there are differences in the way MDLs are determined and reported, the method of determining the MDL for a particular measurement must be known and understood.

b. The MDL is a variable itself and methods for determining MDLs can vary from laboratory to laboratory.

For example, an MDL value could have been calculated as 3X the instrumental limit of detection (IDL) or been based on blank variability for a sorbent-based method. It could also mean that a particular peak or ion in a chromatogram was not seen at all, or that a peak was observed but the concentration was lower than the mean blank, or that the value was below a certain percentile of the distribution of possible blank values. In each case,

the value could be treated differently because their reliability is different, and we may want to substitute them using different approaches.

c. The MDL depends not only on the specific sampling and analysis methods but also on the sample duration, and a number of other factors that may be related to the specific operator of the site and laboratory.

d. The MDL can vary among laboratories even if they follow the same written procedures. Furthermore, there are biases among laboratories.

e. The MDL can vary within a single laboratory because more than one analyst may be involved or for other reasons.

f. The MDL is not a specific and constant value, even for a single laboratory consistently using a standard sampling/analytical methodology. The MDL may vary from day to day and from sample to sample.

Although it would be ideal if each reported measurement had an associated MDL value, it would be extremely cumbersome and expensive to determine it.

The consequence of all these issues is that , before undertaking significant analyses, EPA should understand what each reported MDL value actually means, how it was calculated, and that the replacement approach should include only like-estimates of the MDL. Otherwise there will be additional and inappropriately added variance to the data set, which will be reflected in the estimation of means and variances.

3.4.3.3 Alternate Statistical Approaches for Data Below the MDL There are a number of approaches available to obtain robust estimates of the parameters of the distribution of concentrations when a significant fraction of the observations is below the MDL. The Subcommittee suggested two types of approaches for addressing this issue. Approaches that apply at the front end, that is at the time the laboratory analyses are being conducted, and approaches that apply at the back end, that is once the data base has been assembled.

Front-End Approaches There are a number of approaches that may offer improved estimates of the parameters of the distribution-one at the front end, at the time the laboratory analyses are being conducted, and the other at the back end, once the data base has been assembled. Some examples are provided below.

Sampling and/or analysis methods may be modified to increase the sensitivity of the measurement, so that the percentage of the concentration data that falls below the MDL is very small. Such modifications can be done at the point of sampling, chemical analysis, and/or data reporting.

Sample duration can be lengthened, so that more analyte is collected. There are both practical limitations to this approach and potential technical problems due to sampling artifacts (e.g., decomposition of more reactive analytes during prolonged sampling times).

The method of "standard additions" can be used at the point of analysis. In this method, a known amount of analyte (standard) is added to a sample matrix. The idea is that the standard's response, as measured by the instrument, will increase (adjusted for any dilution and matrix effects) proportionally to the concentration of the analyte in the sample. Typically this is done at three different concentration levels, usually consistent with points on a calibration curve. If the slope of the standard additions curve parallels that of the calibration curve, then the net difference between the curves at the y-intercept is the concentration of the analyte of interest. In essence, the added amount produces a measurable value above the MDL, and the known amount is subtracted, giving the remaining value of the analyte. As long as the resulting value is greater than the uncertainty in the calibration curve, a reasonable estimate for the original non-detectable concentration can be made. The drawback to this method is that it is very labor intensive, requiring at least two, and preferably three additional runs for each sample with non-detectable concentrations. This approach is very difficult to apply to canister sampling. Where there are a significant number of non-detectable concentrations, this process can substantially increase laboratory analytical costs.

Another front-end approach involves data reporting. If both the confidence about the MDL estimate and the value are reported, then this uncertainty can be incorporated in any comparison with a corresponding model estimate of concentration.

Just because an analytical result is below the MDL does not mean that the laboratory has not been able to measure a value, but rather that the measurement has less reliability than others that are above the MDL. When background concentrations are not an issue (i.e., if the method is sorbent-based and there is a background of the analyte, then the sample MDL is determined differently) MDL's are determined by running a low-level standard many times (e.g., 20 runs), and determining the variability in terms of its standard deviation about the measured mean. Three times the standard deviation (3-sigma) is added to the y-intercept of the calibration curve to determine the MDL. This results in a 99% confidence level that the data points above the MDL are quantifiable. At 2-sigma, the confidence is about 95%; at 1-sigma, about 65%. Thus, values measured below the MDL could be reported along with its corresponding sigma value.

Measurements reported as below the 1-sigma level would be considered to be in the noise, and not reportable. Those using the data could then determine what confidence interval is appropriate, and select values based on the desired level. Several Subcommittee members stated that it is more useful to have laboratories report all data with associated uncertainties than to have the laboratories censor the data. This approach is much less labor intensive than the "standard additions" approach, in that additional laboratory analyses are not necessary.

Back-End Approaches Once a measurement result has been reported as below the MDL, there is a question of to how to handle this result in subsequent mathematical and statistical calculations. Gilbert and Kinnison (1981) reviewed some of the more popular methods of dealing with this problem. Other methods have been proposed (see Schmoyer et al. 1996).

The choice of back-end method depends upon what assumptions are consistent with the form of the underlying statistical distribution of the measurement data. For example, it is not appropriate to assume that a set of measurement results at any sampling location should fit a statistical distribution when these observations may be more likely to reflect a time- or spatial series. Thus, that which appears as a useful statistical model from an observed data distribution, e.g. normal, lognormal, Weibull models, may in reality result from a non-random process. This frequently compromises the statistical justification behind a method for dealing with observations reported as below the MDL.

The convention of using half the MDL replacement is simple and based in an acceptable rationale. If there is a set of measurements with unknown concentrations, but it is known these values are between true 0 and an upper limit, and it is further assumed that the distribution of those values between true 0 and the upper limit is normal and the MDL values in the data set are an unbiased sample of that distribution, then the best estimator of the mean of the distribution is the mean of the MDL estimates, which approximates half the MDL. This approach is fairly robust in many cases, but it becomes less so as the proportion of values below the MDL increases and/or the distribution of the data deviates strongly from normalcy. Then, the method chosen (i.e., the one that will provide the most stable estimate of the mean) depends on what we know about the distribution of the data, the percentage of values below the MDL, and the eventual uses of the data. It is also important to consider if the additional computational effort is cost effective.

The key questions are:

How much does the selected substitution affect the measure of central tendency and variance?

Is the effect important given a pre-selected criterion for a good fit between the mean of the measurement and the modeled estimate of concentration?

As an example of the second question, the effect of the replacement method on the estimated distribution parameters may not be important if the criterion for good match is one order of magnitude. However it could be very important if the criterion for good fit is a factor of 2.

When comparing sets of data pair wise for the purposes of determining if they are similar or not making use of values below the MDL becomes even more complex, particularly when the total number of pair values is low, and the proportion of data below the MDL is high. Examples of these types of pair comparison are (1) the same compound measured by two different methods or (2) measured concentrations compared with model estimates. Example (2) is closest to the comparison of the parameters of the distribution of measurements to a single model estimate of average concentration presumably without variability. In this case, the answer can be succinctly described as the replacement method that provides the most stable estimate of the measure of central tendency of the distribution of measurements.

There are a number of other approaches available to make use of measurements below the MDL, including replacement by MDL divided by the square root of 2, Monte Carlo methods, random selection below the MDL (if the distribution of the data is known), and maximum likelihood methods. Discussion of these
various methods and the situations where they might apply can be found in a number of sources. Some user-friendly description with direct application to environmental data can be found, for example, in Gilbert, 1987.

A particularly attractive alternative approach is to utilize known information to estimate the unknown information. Given the percentage of values below the MDL, the mean and standard deviation for the values above the MDL, the underlying distribution (e.g., lognormal), the mean and standard deviation for the entire distribution could be estimated. Or, one could use this technique to create a simple look-up table which estimates the below-MDL values based on the percentage of non-detects, for example:

If 50% of data are below the MDL, the average of those values is 50% of the MDL;

if 25% of the data are below the MDL, the average of those values is 75% of theMDL;

if 90% of the data are below the MDL, the average of those values is 10% of the MDL; etc.

However, these methods can only be applied if the shape of the distribution is known. If, for example, the mode of the distribution is higher than the MDL value, then other approaches may be needed, probably a maximum likelihood approach (see, for example, the Cohen method in Gilbert, 1987, pg 182-183).

Finally, for purposes of Model-to-Monitor comparisons, it may be more useful to set low-level model predictions to the monitored value reported as below the MDL. This might well minimize the impact of the choice of convention on dealing with measurements reported as below the MDL on comparisons between model predictions and monitor data.

GLOSSARY

Air Toxics Program (ATP)

Air toxics 188 hazardous air pollutants regulated under the Clean Air Act (ASPEN)

Clean Air Act (CAA)

Data Quality Objectives (DQO)

Environmental Protection Agency (EPA)

Gas chromatography/mass spectrometry (GC/MS)

Generally Achievable Control Technology (GACT)

Government Performance and Results Act (GPRA)

Hazardous air pollutants (HAPs)

High pressure liquid chromatography/mass spectrometry (HPLC/MS)

Inorganic method (IO)

Instrumental limit of detection (IDL)

(MATES-II)

Maximum Achievable Control Technology (MACT),

Method detection limit (MDL).

National Air Toxics Assessment (NATA).

National Toxics Inventory (NTI)

Office of Air Quality, Planning, and Standards (OAQPS)

Office of Solid Waste and Emergency Response OSWER

Polynuclear aromatic hydrocarbons PAHs.

(PAMS)

Persistent Bioaccumulative Toxics (PBT)

Science Advisory Board's (SAB's)

Total Maximum Daily Load (TMDL)

Toxic Organic method (TO)

Total Risk Integrated Methodology (TRIM)

Multi-media fate within the TRIM package (TRIM.FaTE)

Urban Air Toxics Strategy

Volatile Organic Compounds VOC

REFERENCES

Gilbert, Richard O. (1987) Estimating the Mean and Variance from Censored Data Sets, In: Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold Company, New York, NY., pp 177 - 185

Gilbert, R.O., and R. R. Kinnison, 1981, ìStatistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported or Less-Than Values, Health Physics Vol. 40, pp. 377-390.

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr, 1996, “Difficulties with the lognormal model in mean estimation and testing.” Environmental and Ecological Statistics, Vol. 3, No. 1, pp. 81-97.

Gilbert, R. O., and J. C. Simpson, 1990, “Statistical sampling and Analysis Issues and Needs for Testing Attainment of Background-Based Cleanup Standards at Superfund Sites,” In Proceedings of The Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site: Protocols, Tools, and Research Needs, Arlington VA, February 21-22.

National Environmental Trust, “Up in Smoke: Industry Lobbyists' Big Lie about Barbecues and the New Clean Air Act Standards”, (Table 1), May 1997

USEPA, 1994, Methods for Evaluating the Attainment of Cleanup Standards. Volume 3: Reference-Based Standards For Soils and Solid Media, Washington D.C., EPA 230-R-94-004.

U.S. Nuclear Regulatory Commission, 1995, A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys. NUREG

APPENDIX A:Exposure Measurement Issues

Exposure measurement error not only includes any errors resulting from the measurement instrument, but also considers the error in assigning an individual's exposure based on instruments some distance away (spatial) from each individual in the study population. Measurement errors may consist of “classical” error, which causes bias in measures of association in most situations, and “Berkson” error, which causes little or no bias. Berkson error occurs when the expectation of the measured value is not the true value but near the average of the true values. There are three components of exposure measurement error:

1. the individual exposure and average of personal exposures,
2. the average personal exposure and true ambient levels, and
3. the measured ambient level and true ambient level.

“True” exposures cannot easily be measured. The major Berkson error component is the difference between an individual's actual exposure to a particular pollutant and the average individual exposures of everyone in a geographical area of interest. The average individual exposures will not be known in NATA, but instead the ambient levels will be measured by one or a few monitors in an area. The difference between the monitor measurements and the average personal exposure is the remaining error and is more of the “classical” error that is likely to introduce bias in the risk estimate.

This remaining error can be further decomposed into: the difference between the average personal exposure and the true ambient level, and the difference between the true ambient level and the measured ambient level. The difference between the true and measured ambient level probably would not introduce bias if the average measurement from available monitors is an unbiased estimate of the true, spatially averaged ambient level. This leaves the difference between average personal and ambient levels as the most likely cause of bias. In NATA the average personal exposure will be modeled from the measured ambient level data. An approach that can be used to correct for such biases is to use regression calibration which uses data on both the error-prone daily neighborhood and fixed-site ambient level measurements and personal exposure measurements for some persons on the same days. Such data can be used to calibrate, that is, adjust, the ambient exposure measures by estimating from a regression model the change in average personal exposures corresponding to a unit change in ambient levels. Once the calibration factor is known, the estimated change in risk per unit change in

ambient levels can be corrected so that they apply to changes in personal exposures. Regression calibrations can be obtained from TEAM, PTEAM, NHEXAS, EXPOLIS and THEES data where personal and ambient levels were measured for a number of the HAPS.

APPENDIX B: Summary of Elements of the EPA Quality System and an Introduction to the Data Quality Objectives Process

The Agency's quality **policy** is consistent with ANSI/ASQC E-4 and is defined in EPA Order 5360.1 CHG 1 (1998), the Quality Manual and the organizational components designed for policy implementation as described by the Agency's **Quality System** (EPA QA/G-0). The quality system provides the framework for planning, implementing, and assessing work performed by the organization for carrying out required quality assurance and quality control.

EPA has a comprehensive system of tools for managing its data collection and use activities to assure data quality. The **management tools** used in the organizational level of the EPA Quality System include Quality Management Plans and Management System Reviews. The **technical tools** used in the project level of the EPA Quality System include the Data Quality Objectives Process, Quality Assurance Project Plans, Standard Operating Procedures, Technical Assessments, and Data Quality Assessment.

At the management level, the **Quality System** requires that organizations prepare **Quality Management Plan** (QMP). The QMP provides an overview of responsibilities and lines of authority with regards to quality issues within an organization. Therefore, not only does ETV have a QMP, but the verification partners and subcontractors are required to develop and implement their own QMPs. The ETV program calls these documents **Quality and Management Plans.**

Organizations with **QMPs** review their own performance and develop **Quality Assurance Annual Report and Work Plans** (QAARWP) that provide information on the previous yearís QA/QC activities and those planned for the current year. The QAARWP functions as an important management tool at the organizational level as well as at the Agency-wide level when QAARWP supplied information is compiled across organizations.

At longer multi-year intervals EPA conducts periodic **Management System Reviews** for organizations. An **MSR** consists of a site visit; a draft report that details findings and recommended corrective actions, consideration of the reviewed organization's formal response to the draft report and the authoring of a final report.

At the project level, the data life cycle of planning, implementation and assessment becomes important. The data life cycle begins with systematic planning. EPA recommends that this required planning be conducted using the **Data Quality Objectives (DQO) Process**. The DQO process is a strategic planning approach based on the scientific method that is used to prepare for a data collection activity. It provides a systematic procedure for defining the criteria that a data collection design should satisfy, including when to collect samples, where to collect samples, the tolerable level of decision errors for the study, and how many samples to collect.

EPA has prepared *Guidance for the Data Quality Objectives Process* (QA/G-4). This guidance document applies to projects where the objective of the study is to collect environmental data in support of an Agency program, and, the results of the study will be used to make a specific decision. DQOs are qualitative and quantitative statments that clarify study objective(s), define the most appropriate type of data to collect, determine the most appropriate conditions from which to collect the data, and specify tolerable limits on the decision errors which will be used as the basis for establishing the quantity and quality of data needed to support the decision. The QA/G-4 provides guidance on using a systematic planning process to develop DQOs; it is based on a graded approach.

Briefly, the seven steps in the DQO process are:

1. State the problem
2. Identify the decision
3. Identify the inputs to the decision
4. Define the study boundaries
5. Develop a decision rule
6. Specifiy tolerable limits on decision errors
7. Optimize the design

The **Quality Assurance Project Plan (QAPP)** is the principal output of the **DQO** process and is the project-specific blueprint for obtaining data appropriate for decision-making. The QAPP translates the DQOs into performance specifications and QA/QC procedures for the data collectors. **QAPPs** provide a second level of assurance that the test will be performed in a matter to generated objective and useful information of known quality.

The final step in the data life cycle is the **Data Quality Assessment (DQA)** which determines whether the acquired data meet the assumptions and objectives of the systematic planning process that resulted in their collection. In other words, the DQA determines whether the data are usable because they are of the quantity and quality required to support Agency decisions.